Electric Power Systems Research 117 (2014) 163–171

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Development of distinct load signatures for higher efficiency of NILM algorithms

Aggelos S. Bouhouras*, Apostolos N. Milioudis, Dimitris P. Labridis

Power Systems Laboratory, Dept. of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

ARTICLE INFO

Article history:
Received 13 July 2013
Received in revised form 30 May 2014
Accepted 20 August 2014

Keywords:
Load signatures
NILM algorithms
Load identification
Smart grids
LV measurements
Appliance categorization

ABSTRACT

A simple and novel methodology for the configuration of robust and distinct load signatures is presented in this paper. A technique based on a spectral analysis of the current waveform of each load is proposed, leading to load signatures consisting of three distinct spectral coefficients. Data required for each load signature come from only one period of the current waveform, under a sampling frequency at the order of kHz. A simple identification algorithm has also been developed in order to estimate the signatures adequacy. Each appliance has been measured and classified as an individual load, using a simple identification procedure. Data obtained from two different measurement sets show that the proposed technique provides unique data formations for each appliance tested by developing an efficient recognition procedure with high levels of identification rate. Due to the simplicity of both the database configuration and the identification procedure, the proposed methodology could be considered appropriate for online monitoring and real time smart grid applications.

1. Introduction

The implementation of Smart grid applications has been evolving to a necessity during the last decade, mainly due to the transition of the way distribution networks (DNs) are managed from the traditional conventional way to a more sophisticated and intelligent one. The dynamic behavior of the DNs is enhanced by the penetration of distribution generation (DG) [1] and by a large number of automation applications [2–4] for the so called real time operation of the network. Within this context, smart home applications are becoming quite representative small-scale implementations toward a more efficient management of the network loads. Such applications are based on online monitoring and control of the low voltage (LV) loads and offer the successful operational status identification of each LV load.

The core of these applications is their capability to develop appropriate data forms describing the behavior of each load in a unique and representative way. Traditionally, the latter has been addressed either by installing sensors on every appliance or an intermediate monitoring system in order to record its operation [5]. This intrusive load monitoring method is considered inconvenient, due to its high cost for large scale implementations. A more simple methodology, namely the non-intrusive load monitoring (NILM), has been proposed at the early 1990s [6]. NILM has the advantage of requiring only a single power meter installed at the main feeding panel of each premise, in order to monitor and identify the loads. Although this approach leads also to lower implementation cost, its prime challenge so far consists in the way the load identification is provided efficiently from aggregated signals. NILM algorithms rely on the utilization of the electrical and functional characteristics of the loads in order to create distinct and robust data fingerprints, called from so on load signatures (LS). The higher the uniqueness of these LSs, the easier the identification procedure. This led to a lot of relevant research during the last decade [7–11], since LSs constitute the key for successful load recognition. New load identification methods based on the NILM approach have been also proposed [12–16]. In these approaches, several load features such as active and reactive power, harmonic distortion, transient behavior, and even voltage distortion have been used in order to create a unique LS.

In this work, a novel analysis regarding the development of distinct LSs by using a new and simple methodology is presented. LSs formation is based on a spectral distribution analysis of the load current waveform, with the help of which a number of coefficients is computed. The proposed algorithm depends on simple mathematical equations applied on a limited number of measurement data and thus it benefits in computational time. The required initial data for each appliance signature formulation is only one period

* Corresponding author at: Power Systems Laboratory, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, P.O. Box 486, 54124 Thessaloniki, Greece. Tel.: +30 2310996335; fax: +30 2310996302.
E-mail addresses: abouchou@auth.gr, ag_bou@yahoo.gr (A.S. Bouhouras).

http://dx.doi.org/10.1016/j.epsr.2014.08.015
0378-7796/

(i.e. 20 ms) of the current waveform, using a sampling frequency of 10.24 kHz. This was just the default sampling frequency of the used measurement device. The signal is afterwards transformed using a fast Fourier transform (FFT) and the whole spectrum is accordingly divided to a number of non-overlapping frequency bands. Finally, three special coefficients, namely the Shannon entropy (SE), the Renyi entropy (RE) and the spectral band energy (SBE), are computed and their values are properly used for each LS creation. Results indicate that for the individual appliance case the identification procedure behaves quite efficiently. The proposed technique could therefore be considered suitable not only for standalone load identification but also for event based load recognition.

The paper is organized as follows: in Section 2, essential information regarding the collection of the required data for the examined loads is given. In Section 3, the proposed technique concerning the formulation of the LSs along with the corresponding identification algorithm for the examination of the effectiveness of the formulated signatures is presented. In Section 4, results of the implemented simulations regarding the identification of each examined appliance considered as an individual load are illustrated. Section 5 is finally devoted to conclusions.

2. Input data

2.1. Examined appliances

In the present analysis, representative residential LV appliances that are presented in Table 1 are examined. Various function modes have been also taken into account for some appliances, so long as appliances shown in Table 1 refer to loads having different electrical characteristics. All possible electrical behaviors, e.g. purely resistive, combination of resistive/capacitive and resistive/inductive, and even harmonic polluting ones, have been therefore examined in order to evaluate the applicability of the proposed technique.

2.2. Measurement implementation

Two basic measurement sets have been implemented, in order to map the recorded signal to a suitable LS of each appliance. All appliances have been measured in laboratory conditions, while some of them have been also measured in field conditions. Measurements have been accomplished in different days and times, in order to examine whether results could possibly had been influences by the slightly different ambient conditions. *Fluke 1760 – Three Phase Power Quality Recorder*, a device fully compliant with IEC 61000-4-30 Class-A standard, calibrated by Hellenic Institute of Metrology (EIM) just before the measurements has been used for all measurements. Results showed that the active power uncertainties were much less than the corresponding divergence of all appliances. The current waveform of each appliance has been monitored and recorded under a sampling frequency of 10.24 kHz, which was the default sampling rate of Fluke 1760. Aiming to acquire the most possible reliable results, measurement of the grid background, i.e. measurement at the feeding point of the installation by having all appliance turned off, has preceded the load measurements. Finally, results of these two measurement sets were compared in order to define the incoming grid noise, which for the presented measurements has been found negligible.

Table 1
LV appliances investigated.

Appliance	Consumed active power (W)
(1) Air conditioner with inverter	1080
(2) Air conditioner without inverter	1600
(3) Coffee machine	1000
(4a) Hair dryer-full start mode	1200
(4b) Hair dryer-half start mode	760
(5a) Heater-full resistance mode	2000
(5b) Heater-half resistance mode	1000
(5c) Heater-no resistance mode	15
(6) Home theater	14
(7) Electric iron	2100
(8) Laptop-operational mode with battery only	45
(9) Laptop	70
(10) Refrigerator	140
(11) Washing machine	2400
(12) Halogen lights	90
(13) Led lights	9
(14) Luminaire	15

3. Proposed methodology

3.1. Database formulation

Although the combination of several characteristics of a load is generally expected to increase the robustness of its signature, in this work the simplicity of the whole identification procedure has been considered as a priority. Therefore, the needed data for the development of each LS obtained from only one period of the current waveform, using the sampling frequency of 10.24 kHz consisted approximately of 205 instant current values. These values have been processed by the following procedure:

(1) Initially the group A of n examined appliances is considered, $a = 1, \ldots, n$ (in this work $n = 17$ since the different function modes of some appliances are treated as different appliances).

Let us consider a current waveform $f(t)$ in the time domain and $F(\omega)$ its image in the frequency domain. The numerical formulation implementing N total points sampling uses a discretization in the frequency domain with spacing $\Delta\omega$ and time steps of Δt in the time domain, as shown in the next two equations:

$$f_n = f(n\Delta t), \quad n = 0,\ 1,\ \ldots,\ N-1 \tag{1}$$

$$F_k = F(jk\Delta\omega), \quad k = 0,\ 1,\ \ldots,\ N-1 \tag{2}$$

The following equations relate the total observation time T, the total number of the sampling points N and the maximum corresponding frequency Ω with the discretization spacing in time domain (i.e. Δt) and frequency domain (i.e. $\Delta\omega$) [17].

$$\Delta t = \frac{T}{N} \tag{3}$$

$$\Delta\omega = \frac{2\Omega}{N} \tag{4}$$

Moreover, the frequency domain image of the measured current input signal is separated into several frequency bands i, where f_{i1} and f_{i2} are the minimum and maximum frequency of the ith band respectively.

$$F_{ki} = \{F_k : k \in [f_{i1}, f_{i2}]\}$$

(2) Consider as the input signal a vector of 205 values of the current waveform during the steady state for appliance α $[I_{ti}^{a}]$ with $i = 1, \ldots, 205$ and t referring to the time domain.

(3) Apply FFT to the input signal, in order to create the current vector in the frequency domain: $[I_{ti}^{a}] \xrightarrow{\text{FFT}} [I_{fi}^{a}]$, with f referring to the frequency domain.

Table 2
Seven logarithmically spaced bands.

Band *Bj*	Lower bound *lj*	Upper band *uj*
1	0	62.5
2	62.5	125
3	125	250
4	250	500
5	500	1000
6	1000	2000
7	2000	5512

(4) Define the number and the bandwidth of the non-overlapping frequency bands *B*s. Seven non-overlapping logarithmically spaced bands were considered initially. In Table 2 each of these bands B_j with $j = 1, \ldots, 7$ is shown, along with its corresponding upper and lower bounds.

$$B_j = b_{l_j u_j} \quad \text{with} \quad \left\{ \begin{array}{l} l_j = \text{lower bound for } j \text{ band} \\ u_j = \text{upper bound for } j \text{ band} \end{array} \right\}$$

(5) Compute three specific coefficients related to the spectral distribution of the input signal. These coefficients [18] are as follows:

- *Shannon entropy* (SE) coefficient. The SE of a signal is a measure of the spectral distribution of the signal. SE is defined as:

$$\mathrm{SE}_j^a = \sum_{l_j}^{u_j} |I_{fi}^a| \cdot (\log_2 |I_{fi}^a|) \tag{5}$$

- *Renyi entropy* (RE) coefficient. The RE of a signal is also a measure of its spectral distribution. RE is defined as:

$$\mathrm{RE}_j^a = \frac{1}{1-r} \cdot \log \left(\sum_{l_j}^{u_j} |I_{fi}^a|^r \right) \quad \text{with} \quad r = 2 \tag{6}$$

The value order of 2 regarding the r constant was chosen due to no further improvement to the identification rate for higher values ($r > 2$).

- *Spectral band energy* (SBE) coefficient. The SBE is the energy in the frequency bands normalized by the energy of the whole spectrum. SBE is defined as:

$$\mathrm{SBE}_j^a = \frac{\sum_{l_j}^{u_j} |I_{fi}^a|^2}{\sum_{l_1}^{u_7} |I_{fi}^a|^2} \tag{7}$$

(6) Define the value of the SE coefficient for the first band B_1.

Based on a thorough process of the computed values for SE, it has been observed that its magnitude for the first band could be either positive or negative. This observation, explained mainly by the nominal current of the loads, helped to apply a first heuristic rule in order to group the appliances into two subsets. Appliances having small rated power were assigned a negative first band coefficient, while appliances having higher rated power a positive one. By using this rule, an initial categorization regarding the rated power of the appliances is being implemented. Thus, the first band B_1 takes into account the nominal load current, transforming this rule to a power based categorization criterion. The implementation of this rule is presented as follows:

$$\text{is} \quad \mathrm{SE}_1^a < 0? \rightarrow \left\{ \begin{array}{l} \text{if YES then } \mathrm{SE}_1'^a = 1 \\ \text{if NO then } \mathrm{SE}_1'^a = -1 \end{array} \right\} \tag{8}$$

(7) Define whether the magnitude of the value of SBE coefficient for the fourth band B_4 is lower than a predefined threshold T, which was considered equal to 0.01.

Based on the coefficient's values for the fourth band it was observed that an additional criterion for load categorization could be developed. Threshold T was determined arbitrary based on the observed values of the respective coefficient. Different thresholds are expected to result in different appliance grouping, which in turn affects only the computational time of the algorithm and not its load identification performance. The heuristic rules for load categorization could be unlimited, while the benefit relies on breaking up the solution space in smaller parts in order to fasten the identification procedure. The latter means that the algorithm initiates its matching procedure by targeting in a specific subgroup or subgroups of the database and thus implementing the load identification in a shorter time period. This second heuristic rule was developed in order to improve the discretization of the subsets and increase their number. The implementation of this rule is presented as follows:

$$\text{is} \quad \mathrm{SBE}_4^a < \text{threshold } T? \rightarrow \left\{ \begin{array}{l} \text{if YES then } \mathrm{SBE}_4'^a = 1 \\ \text{if NO then } \mathrm{SBE}_4'^a = -1 \end{array} \right\} \tag{9}$$

The selection of heuristic rules involves some level of arbitrariness; however, these rules do not affect the efficiency of the proposed identification procedure, since they are just expected to provide the capability for grouping the appliances in subsets regardless the number of the loads existing in the database. Therefore, a version of the proposed algorithm with no heuristics provides the same results but the identification part could be more time consuming (the latter is expected to be noticed only for large data bases). A justification regarding the selection of the specific heuristic rules relies on the following: the value of the first band of the SE coefficient is quite representative of the consumed active power of each appliance, since it refers to the first utilized band containing the fundamental frequency (since the bandwidth of the first band is 0–62.5 Hz). Moreover, the SBE coefficient has been chosen due to its mathematical expression; it provides information for each appliance regarding its harmonic content. Therefore, this coefficient's value corresponds to harmonic content level that exists. Also, the value of SBE coefficient it was selected for the fourth band because it could provide a better grouping criterion, in comparison to the other six bands. The two heuristic rules constitute categorization criterions regardless the loads examined, since in this work the database of the utilized appliances is considered as representative of the loads that the proposed algorithm is designed to deal with, which are residential LV appliances. However, any new appliances could be added to the database and would just be assigned to one of the already created groups by using these rules.

(8) The two heuristics rules defined in steps 6 and 7 contribute in developing an initial format of each LS by the sense of structuring a simple type of label for each appliance. This kind of label, called the first part of the load signature (be it LS1), is defined as follows:

$$[\mathrm{LS1}]^a = [\ \mathrm{SE}_1'^a \quad \mathrm{SBE}_4'^a\ ] \tag{10}$$

Based on (10), three subsets G_A, G_B and G_C of the initial considered group A of appliances are originally formed

as follows:

$$\left\{ \begin{array}{l} G_A \text{ with} w \text{ appliances, } w \in A \quad \text{if} \quad \left\{ \begin{array}{l} \mathrm{SE}_1^a > 0 \\ \mathrm{SBE}_4^a < T \end{array} \right\} \\ G_B \text{ with} z \text{ appliances, } z \in A \quad \text{if} \quad \left\{ \begin{array}{l} \mathrm{SE}_1^a < 0 \\ \mathrm{SBE}_4^a > T \end{array} \right\} \\ G_C \text{ with} v \text{ appliances, } v \in A \quad \text{if} \quad \left(\begin{array}{l} \left\{ \begin{array}{l} \mathrm{SE}_1^a < 0 \\ \mathrm{SBE}_4^a < T \end{array} \right\} \\ \text{or} \\ \left\{ \begin{array}{l} \mathrm{SE}_1^a < 0 \\ \mathrm{SBE}_4^a < T \end{array} \right\} \end{array} \right) \end{array} \right\}$$

In each of these three subsets, appliances sharing the same first part of signature LS1 are included. For reasons of simplicity and convenience a two digit label is assigned to every subset, i.e. G'_A, G'_B and G'_C, in order to facilitate the assignment of each appliance to the respective subset it belongs. This constitutes a transformation of the coefficients' values to a binary string, given by

$$\left\{ \begin{array}{l} G_A \rightarrow [G'_{\mathrm{A}}] = [1, 1] \\ G_B \rightarrow [G'_{\mathrm{B}}] = [-1, -1] \\ G_C \rightarrow [G'_{\mathrm{C}}] = [1, -1] \text{ or } [G'_{\mathrm{C}}] = [-1, 1] \end{array} \right\} \tag{11}$$

The importance of step #8 relies on the disaggregation of the appliances of the database into three subsets and it results to shorten the computational time of the identification procedure. Its benefit consists in targeting directly to specific smaller subgroup, when aiming to match an unknown appliance (be it $X1$) with the respective existing one in the original database. Such choice could improve significantly the efficiency of the whole procedure, especially for large scale applications having numerous loads.

(9) Finally, the second and more crucial part of the load signature (be it $\mathrm{LS2}^\alpha$) is structured as follows:

$$\mathrm{LS2}^a = \left\{ \begin{array}{ll} [\mathrm{RE}]_j^a & [d_{\mathrm{RE}}]_k^a \\ [\mathrm{SE}]_j^a & [d_{\mathrm{SE}}]_k^a \end{array} \right\} \tag{12}$$

The first column of $\mathrm{LS2}^a$ consists of the coefficients values that have been defined in step #5 and computed for the respective bands. The second column concerns the declination vectors for the corresponding coefficients, when the seven values that have been computed for the seven sequential bands of Table 2 are formed in a graph. The selection of a declination vector for each appliance relies on the formation of an additional fingerprint of each appliance's behavior in order to strengthen the signature's robustness. In Fig. 1 an example for a coffee machine appliance regarding its declination fingerprint of the RE coefficient is presented. The declination vectors are computed by the following equations:

$$[d_{\mathrm{RE}}]_k^a = \arctan[\mathrm{RE}_{j+1}^a - \mathrm{RE}_j^a] \tag{13}$$

$$[d_{\mathrm{SE}}]_k^a = \arctan[\mathrm{SE}_{j+1}^a - \mathrm{SE}_j^a] \quad \text{for} \quad \left\{ \begin{array}{l} j = 1, \ldots, 7 \\ k = 1, \ldots, 6 \end{array} \right\} \tag{14}$$

All above explained steps of the first part of the algorithm regarding the LSs formulation are graphically presented in Fig. 2.

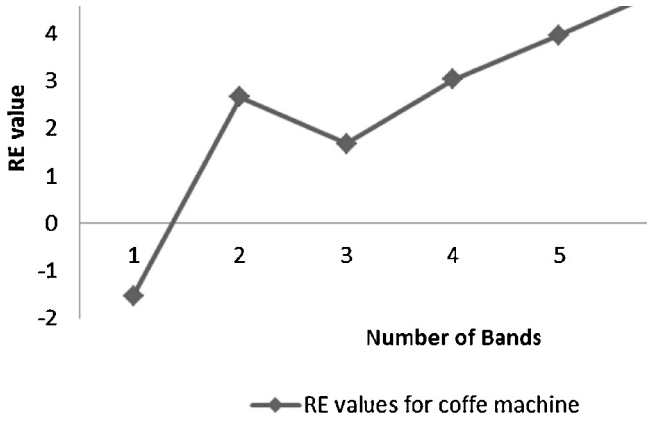

Fig. 1. Graphical form of the RE coefficient values for the coffee machine appliance.

3.2. Identification procedure

In order to examine the effectiveness of the formulated signatures, a simple identification algorithm has been developed. Each appliance has been examined as an individual load by the following procedure:

(1) Apply a simple heuristic rule for the so called event detection. Since only one load is considered each time, the activation of each appliance could be perceptible whenever the instant recorded current magnitude exceeds a predefined threshold TH. In this analysis a threshold equal to 30 mA has been chosen. The reason regarding the latter relies on the measured background before the activation of each load; the magnitude of the recorded instant current varied between 5 and 15 mA. By that sense, the measurements proved that whenever this threshold was exceeded an appliance had been turned on. The feature extraction process is usually triggered after an event is detected. It is often to utilize a predefined threshold regarding the measured current variations [19] in order to decide that a load has been turned on.

(2) For the candidate unknown appliance $X1$ start storing 205 values during its steady state operation (typical a few cycles after the load activation) and under a sampling frequency of 10.24 kHz $\rightarrow [I_t]_i^{X1}$ for $i = 1, \ldots, 205$.

(3) Apply steps 3–9 of the database formulation, in order to formulate the signature of the unknown appliance. The output of this procedure is a two parted signature (as explained in the previous section), [LS1($X1$)] and [LS2($X1$)].

(4) Utilize the first part of the signature in order to specify the subset of the database that the unknown appliance $X1$ could find a match. A simple mathematical formulation for the implementation of this step could be provided by Jaccard index (or Jaccard distance) [20] as follows:

$$J_{\delta 1}([\mathrm{LS1}], [G'_{\mathrm{A}}]) = \frac{|[\mathrm{LS1}] \cup [G'_{\mathrm{A}}]| - |[\mathrm{LS1}] \cap [G'_{\mathrm{A}}]|}{|[\mathrm{LS1}] \cup [G'_{\mathrm{A}}]|} \tag{15}$$

$$J_{\delta 2}([\mathrm{LS1}], [G'_{\mathrm{B}}]) = \frac{|[\mathrm{LS1}] \cup [G'_{\mathrm{B}}]| - |[\mathrm{LS1}] \cap [G'_{\mathrm{B}}]|}{|[\mathrm{LS1}] \cup [G'_{\mathrm{B}}]|} \tag{16}$$

$$J_{\delta 3}([\mathrm{LS1}], [G'_{\mathrm{C}}]) = \frac{|[\mathrm{LS1}] \cup [G'_{\mathrm{C}}]| - |[\mathrm{LS1}] \cap [G'_{\mathrm{C}}]|}{|[\mathrm{LS1}] \cup [G'_{\mathrm{C}}]|} \tag{17}$$

(5) Define J_δ with minimum value (i.e. the minimum value among Eqs. (15)–(17). Then decide that the unknown appliance $X1$ belongs to group G_A, G_B or G_C accordingly under the acceptance that the min J_δ indicates the minimum distance of the unknown appliance $X1$ from the candidate group.

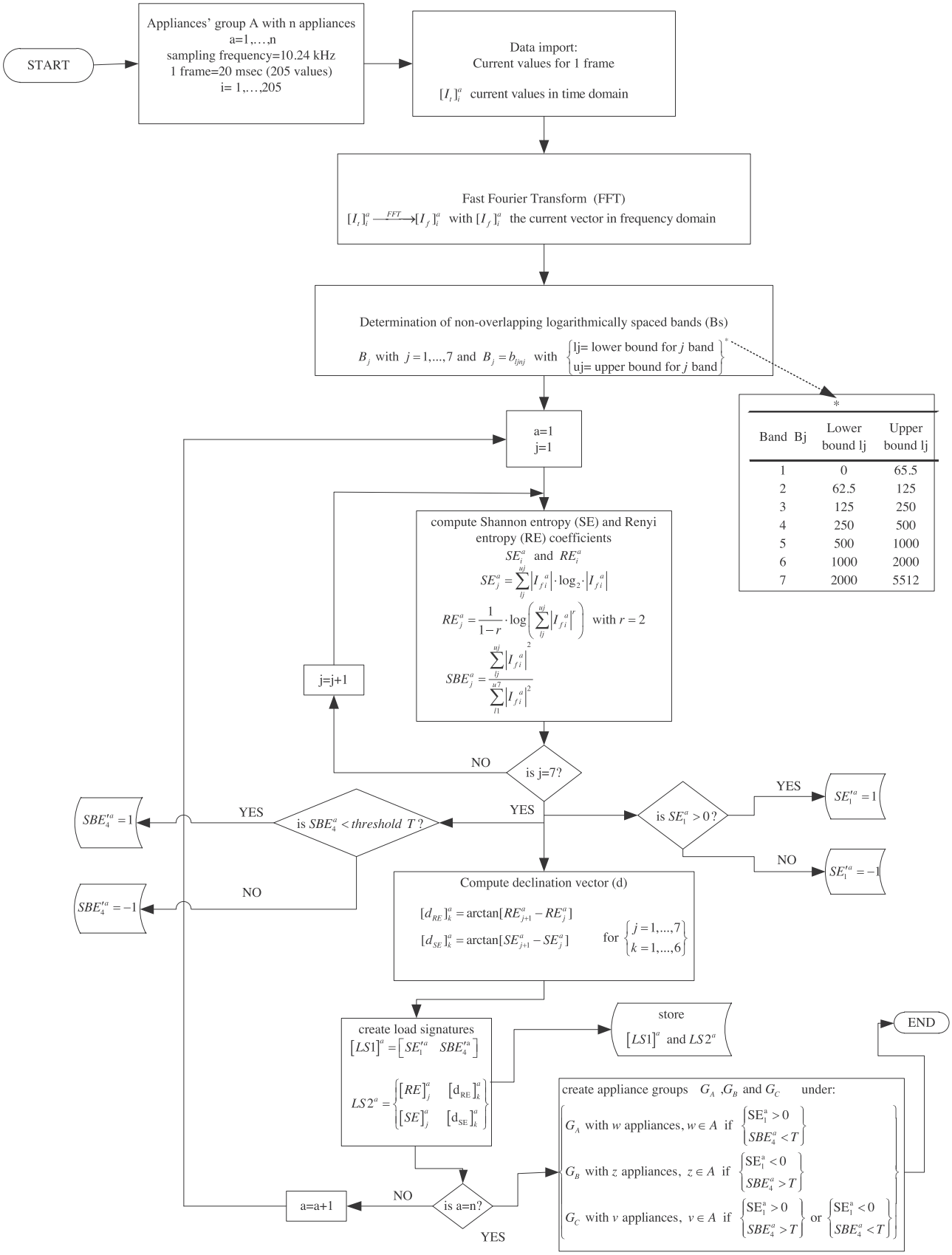

Fig. 2. Flowchart of the database configuration.

(6) Utilize the second part of the signature, in order to match the unknown appliance with the correct one in the already targeted subset. Firstly, an individual residue (IR) for the unknown appliance (one IR value for every feature-coefficient in the second signature part) is defined as follows [21]:

$$IR_{1Y} = \frac{\sum_{Y=1,j=1}^{Y=w\text{ or }z\text{ or }v,\ j=7}(SE_j^{X1} - SE_j^{Y})^2}{\sum_{j=1}^{j=7}(SE_j^{X1})^2} \quad (18)$$

$$IR_{2Y} = \frac{\sum_{Y=1,j=1}^{Y=w\text{ or }z\text{ or }v,\ j=7}(RE_j^{X1} - RE_j^{Y})^2}{\sum_{j=1}^{j=7}(RE_j^{X1})^2} \quad (19)$$

$$IR_{3Y} = \frac{\sum_{Y=1,j=1}^{Y=w\text{ or }z\text{ or }v,\ j=7}\left(d_{SE_j^{X1}} - d_{SE_j^{Y}}\right)^2}{\sum_{j=1}^{j=7}\left(d_{SE_j^{X1}}\right)^2} \quad (20)$$

$$IR_{4Y} = \frac{\sum_{Y=1,j=1}^{Y=w\text{ or }z\text{ or }v,\ j=7}\left(d_{RE_j^{X1}} - d_{RE_j^{Y}}\right)^2}{\sum_{j=1}^{j=7}\left(d_{RE_j^{X1}}\right)^2} \quad (21)$$

In Eqs. (18)–(21), Y is defined as the number of appliances that belong to groups G_A or G_B or G_C. Since the first part of the signature, [LS1($X1$)], is expected to guideline the matching procedure, for every IR the number of the computed differences between the unknown appliance $X1$ and each one of the selected group should be equal to Y.

It is obvious that in Eqs. (18)–(21) the IR of each of the four parts of Eq. (12) of the LS2 is normalized. Afterwards, the four computed IRs for the four different features are formed in a unified residue (UR) as follows:

$$UR_{X1} = \prod_{m=1}^{m=4} IR_{mY} \quad (22)$$

If the [LS1($X1$)] part of the signature (under the step 4 of the identification procedure) indicates the G_A group as the correct subset from which the solution should be examined, then only the SE coefficient should be used. Based on the selected categorization of the appliances it has been observed that G_A subset includes appliances with intense resistive behavior and thus, SE coefficient is considered sufficient for the load identification.

(7) Finally, the minimum value of the UR_{X1} which is defined as the least unified residue (LUR) among the computed Y ones indicates the correct matching of the unknown appliance $X1$ with the correct one in the referred subset. The utilization of LUR provides the identification of the appliance by taking into account all the computed features that constitute the signature simultaneously. The benefit is that the decision index is a real number and thus it will always provide a solution which in most of the times is expected to be unique and correct.

4. Results

4.1. Algorithm implementation

In order to examine the effectiveness of both the developed LSs and the proposed identification procedure, the algorithm has been implemented by using real data from the two performed measurement sets. More specifically, the initial database was formulated by using as fundamental data one random period of the current waveform of each load under steady state conditions. Afterwards, each appliance was tested as an individual load by the sense of investigating whether each load could be identified for individual activation.

The identification part of the procedure involved eighteen input files for each examined appliance. Each of these files consisted of one random but different period of the current waveform under steady state conditions and it was treated as the input file of the candidate unknown appliance that should be identified. So far, each measurement set has been considered as the only available data set for the data base configuration and the loads identification. At a following analysis level, both measurement sets were utilized as a joined data set under the following clarifications: the first set was used for the data base formulation, while the second provided the input files for each appliance. The latter was implemented vice versa as well.

The results derived are presented in Tables 3 and 4 for the two analysis levels respectively. In these tables, the examined cases presented as A, B, C, and D are as follows:

- A: Database formulation and input files from 1st measurement set

Table 3
Results regarding algorithm testing for seven bands.

Appliance	1st analysis level %Successful identification rate		2nd analysis level	
	A	B	C	D
Air conditioner with inverter	100	100	100	100
Air conditioner without inverter	94.44	100	100	100
Coffee machine	88.89	94.44	72.22	100
Hair dryer-full start mode	100	100	100	100
Hair dryer-half start mode	100	100	100	100
Heater-full resistance mode	100	100	100	100
Heater-half resistance mode	100	88.89	100	100
Heater-no resistance mode	100	100	100	100
Home theater	100	100	100	100
Electric iron	88.89	100	100	100
Laptop-battery only	100	100	100	100
Laptop	100	100	100	100
Refrigerator	100	100	100	100
Washing machine	100	83.33	77.78	100
Halogen lights	100	100	100	100
Led lights	100	100	16.67	16.67
Luminaire	100	100	100	100

Table 4
Results regarding algorithm testing for ten bands.

Appliance	1st analysis level %Successful identification rate		2nd analysis level	
	A	B	C	D
Air conditioner with inverter	100	100	100	100
Air conditioner without inverter	100	100	100	100
Coffee machine	100	100	100	100
Hair dryer-full start mode	100	100	100	100
Hair dryer-half start mode	100	100	100	100
Heater-full resistance mode	100	100	100	100
Heater-half resistance mode	100	88.89	88.89	100
Heater-no resistance mode	100	100	100	100
Home theater	100	100	100	100
Electric iron	94.44	100	100	100
Laptop-battery only	100	100	100	100
Laptop	100	100	100	100
Refrigerator	100	100	100	100
Washing machine	100	88.89	88.89	100
Halogen lights	100	100	100	100
Led lights	100	100	61.11	27.78
Luminaire	100	100	100	100

Table 5
Ten logarithmically spaced bands.

Band *Bj*	Lower bound *lj*	Upper band *uj*
1	0	62.5
2	62.5	125
3	125	250
4	250	375
5	375	500
6	500	750
7	750	1000
8	1000	2000
9	2000	3800
10	3800	5512

- B: Database formulation and input files from 2nd measurement set
- C: Database formulation from 1st measurement set and input files from 2nd measurement set
- D: Database formulation from 2nd measurement set and input files from 1st measurement set

The results describe the successful identification rate for each appliance and it is expressed as a % percentage. Based on the proposed approach for the LSs formulation, i.e. on the spectral distribution analysis of the current waveform, the most unique and distinct signatures should be the ones for the harmonic polluting appliances (e.g. air conditioner, home theater, laptop, halogen lights and more). The latter is actually confirmed by the results of Table 3 (i.e. the first two columns with results) especially for the first analysis level. The same observation stands for the second analysis level (i.e. the two last columns of Table 3), for which the database and the input files come from different measurement sets.

At a first analysis level, only for a limited number of appliances the identification rate has not resulted to 100%. These loads concern appliances for which the resistive behavior is dominant and in addition some of them also share almost identical nominal values. By that sense, the identification failures could be considered negligible due to the aforementioned clarifications. In order to increase the efficiency of the methodology and at the same time not to affect its simplicity, the algorithm was tested again for a higher number of bands. It was assumed that a higher number of bands could increase the density of the extracted information by the spectral distribution analysis and thus enhance the robustness of the signatures. Three additional bands are presented in Table 5. As easily observed, two of the three additional bands concern spectral bandwidths between 0 and 1 kHz since at higher frequency levels the harmonic content is expected to be negligible.

The results of the algorithm testing for ten bands are presented in Table 4. It is obvious that the identification rates for the appliances, that were lower than 100% at the seven bands case, are now significantly improved. Although there are few appliances with identification rate lower than 100%, the algorithm fails only twice to identify loads with identical behavior and nominal values. Moreover, appliances with different operational modes could be recognized given that these operational modes have been properly mapped in the database by respective signatures; this case is shown in this work for specific appliances. One more important observation is related to the results at the second analysis level for both cases, i.e. with seven and ten bands respectively. As presented in Tables 3 and 4, the identification rates seem to be the same regardless the utilized data for the database formulation and the testing input files. Only for one appliance, i.e. led lights, the results between the first and the second analysis level are different. The explanation relies on the different magnitude of the current waveform for the two measurement sets. This difference concerns a 33% increase for the instant current values between the two measurements (0.3 A the peak value for the first set and 0.4 for the second). Considering that the peak nominal current value is relatively low (lower than 0.5 A) it is quite difficult to identify the appliance for the case that the signature and the input files come from so different waveforms. Comparing the results between the two test cases it is clear that a higher number of bands contributed in increasing the efficiency of the algorithm. Therefore, the choice of formulating the LSs under a higher number of bands has resulted in more sufficient signatures. It is still under investigation whether an even higher number of bands or different bandwidths of the selected bands could result in more robust and distinct LSs.

As observed in Tables 3 and 4, the proposed algorithm provides good results for the cases where the examined loads do not embed purely resistive behavior. The latter is rational since in this work the proposed coefficients are used in order to extract the maximum possible information about the harmonic content of each respective current waveform. By that sense, for appliances with quasi sinusoidal current waveforms the algorithm is expected to perform well. On the other hand, for appliances with purely resistive behavior, their nominal current taken into account in the first utilized band in SE coefficient, is expected to constitute the crucial criterion for the identification. Also, the proposed algorithm is proved to be very efficient by having an identification rate equal to 100% for the cases where the examined loads do not embed purely resistive behavior. This was expected, since the proposed coefficients are used in order to extract the maximum possible information about the harmonic content of the corresponding current waveform. By that sense, for appliances having quasi sinusoidal current waveforms the algorithm is assumed to perform well. On the other hand, for appliances having purely resistive behavior, their nominal current (in proportion to their nominal active power) is taken into account in the first used band in SE coefficient and is assumed to constitute the crucial criterion for the identification.

The basic advantage of the proposed methodology relies on its simplicity due the limited needed data for the signatures' formulation and the convenience regarding the pattern recognition. In [22] the authors propose a NILM approach that relies on feature extraction by the current and voltage waveforms under the steady state. In this paper seven typical residential appliances are examined, while their LSs are formed based on the active and reactive power that has been measured for each load individually. Although a disaggregation scenario is presented, it should be noted that it actually constitutes an event detection approach. The identification performance is high (i.e. 100%) but no loads with identical active and reactive nominal values have been considered. The benefit of the proposed methodology relies on the exhaustive utilization of the harmonic content of each load under a simple and robust formation that could provide higher efficiency when referring to more loads with similar nominal values but different electrical characteristics. Moreover, in [23] an approach using current harmonic signatures is presented. The disadvantage of this method relies on utilization of one Artificial Neural Network ANN for each appliance, which presumes training and appropriate preprocess of the initial data (not all harmonic signatures utilized because in such case the solution would not be generalized). The algorithm in this paper does not require any training since the signatures are developed only one time provided that there are available 205 instant current values under the steady state. Moreover, only seven appliances are examined and the performance rates are lower than the ones in this analysis. In [24] an analysis close to the present one is proposed. The sampling frequency is almost equal to the one in this paper (10 kHz) and the current waveforms are grouped by the ratio of their fundamental component to the RMS total after a FFT. The

methodology relies on the utilization of a genetic algorithm (GA) that performs efficiently (100%) for non-sinusoidal current waveforms. This methodology lacks in the information extracted by the FFT in the current waveform in comparison to the three coefficients proposed in this paper. Furthermore, the GA is more complex than the matching procedure with the UR in this work. Two additional approaches that use ANNs for their NILM algorithm are presented in [25,19]. In [25] the sampling frequency is 6 kHz and the disaggregation model concerns three case studies with up to five loads operating simultaneously. The utilization of ANN with Gas constitutes this methodology [25] more complex regarding both the development of the algorithm and the training face of the ANNs. In [19], only three loads are examined and despite the utilization of a GA, the identification rates for the energizing case study of the loads is lower than 95%. Furthermore in [26], an evolution of [23] is presented. The authors have increased the sampling from 1 kHz to 20 kHz for a disaggregation case study. The new proposed algorithm could achieve 86.5% device recognition accuracy and the only used attribute is the real power. Although the authors utilize a sampling frequency two times larger than the one in this paper the identification rate is below 90%. In a recent published research [27], load signatures are formulated in a three dimensional space under the steady state. The results prove that the algorithm is efficient enough but the sampling frequency is relatively low. Therefore, in some cases of appliances having almost equal nominal power characteristics, the signatures are not robust enough to provide load identification. Moreover, in this work only one feature (i.e. instantaneous current) is used for the development of the load signatures and this is expected to make the proposed algorithm more simple and time efficient for large scale applications. In [28], the authors also utilize three features of each appliance, active and reactive power and total harmonic distortion (THD), but for the computation of the THD they utilize the harmonic orders up to the ninth which means that for some appliances there could be a lack of information regarding their harmonic content and in turn their signatures would not be as distinct as they could be. Finally in [29] a different approach for a NILM approach with EMF sensors is presented. The method is expected to enhance the efficiency of the NILM platform but the requirement of additional sensors close to the appliances mean that this technique is more complex and hardware demanding in comparison with the one proposed in this work.

5. Conclusion

In this paper, a novel and simple methodology for the development of sufficient load signatures and for the improvement of NILM algorithms efficiency is presented. It is generally accepted that the more unique the signature of a load the more easy to identify its operational status. Furthermore, the efficiency of NILM algorithms relies on both the signatures length, i.e. the size of the data base, and on the complexity of the recognition mechanism. In the analysis presented in this paper, an approach considered to deal with both above issues is presented. Load signatures formulation is based on a spectral distribution analysis of each load current waveform. Three special coefficients are computed for a predefined number of spectral bands and each load signature is formed using two distinct parts. The first part classifies the initial set of the examined appliances into three sub sets, using two heuristic rules. The second part is structured under special coefficient values and constitutes the fundamental tool for each appliance identification.

The results indicate that the proposed methodology provides successful load signatures for individual load activation. Moreover, the analysis showed that the efficiency of the recognition procedure could be improved when the number of the predefined bands is increased. Finally, the methodology has proved to work efficiently even when the data base and the testing input files come from different measurement sets. Conclusively, the paper proposes a novel technique for the configuration of unique load signatures which is expected to enhance the performance of NILM signatures.

Acknowledgements

The research leading to the results described in this paper is part of the European Metrology Research Program (EMRP) ENG04-REG1, which is jointly funded by the EMRP participating countries within EURAMET and the European Union.

References

[1] S. Inglis, G.W. Ault, A. Alarcon-Rodriguez, S.J. Galloway, Multi-objective network planning tool for networks containing high penetrations of DER, in: Proc. UPEC Conference, 2010, pp. 1–6.
[2] I.S. Baxevanos, D.P. Labridis, Implementing multiagent systems technology for power distribution network control and protection management, IEEE Trans. Power Deliv. 22 (2007) 433–443.
[3] A. Dimeas, N. Hatziargyriou, A multiagent system for microgrids, in: Proc. IEEE Power Eng. Soc. General Meeting, vol. 1, 2004, pp. 55–58.
[4] C.P. Nguyen, A.J. Flueck, Agent based restoration with distributed energy storage support in smart grids, IEEE Trans. Smart Grid 3 (2012) 1029–1038.
[5] J. Liang, S. Ng, G. Kendall, J. Cheng, Load signature study – Part I: Basic concept, structure, and methodology, IEEE Trans. Power Deliv. 25 (2010) 551–560.
[6] G.W. Hart, Nonintrusive appliance load monitoring, Proc. IEEE 80 (1992) 1870–1891.
[7] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, P. Armstrong, Power signature analysis, IEEE Power Energy Mag. 1 (2003) 56–63.
[8] Y.H. Lin, M.S. Tsai, A novel feature extraction method for the development of nonintrusive load monitoring system based on BP-ANN, in: Proc. 3CA International Symposium, 2010, pp. 215–218.
[9] H.T. Yang, H.H. Chang, C.L. Lin, Design a neural network for features selection in non-intrusive monitoring of industrial monitoring electrical loads, in: Proc. CSCWD Conference, 2007, pp. 1022–1027.
[10] K.K. Simon, J. Liang, W.M. John, Automatic appliance load signature identification by statistical clustering, in: Proc. APSCOM Conference, 2009.
[11] H.H. Chang, K.L. Chen, Y.P. Tsai, W.J. Lee, A new measurement method for power signatures of non-intrusive demand monitoring and load identification, IEEE Trans. Ind. Appl. 48 (2011) 764–771.
[12] H.H. Chang, C.L. Lin, H.T. Yang, Load recognition for different loads with the same real power and reactive power in non-intrusive load-monitoring system, in: IEEE Proc. CSCWD Conference, 2008, pp. 1122–1127.
[13] W.L. Chan, A.T.P. So, L.L. Lai, Harmonics load signature recognition by wavelets transforms, in: IEEE Int. Conf. on DRPT, 2000, pp. 666–671.
[14] H.-H. Chang, C.-L. Lin, J.-K. Lee, Load identification in nonintrusive load monitoring using steady-state and turn-on transient energy algorithms, in: IEEE Int. Conf. CSCWD, 2010, pp. 27–32.
[15] L.I. Eguiluz, M. Manana, J.C. Lavandero, Voltage distortion influence on current signatures in nonlinear loads, in: IEEE Power Engineering Society Summer Meeting, vol. 2, 2010, pp. 1165–1170.
[16] Z. Wang, G. Zheng, Residential appliances identification and monitoring by a nonintrusive method, IEEE Trans. Smart Grid 3 (2012) 80–92.
[17] P. Moreno, A. Ramirez, Implementation of the numerical Laplace transform: a review task force on frequency domain methods for EMT studies, Working Group on modeling and analysis of system transients using digital simulation, General Systems Subcommittee, IEEE Power Engineering Society, IEEE Trans. Power Deliv. 23 (2008) 2599–2609.
[18] A. Ramalingam, S. Krishnan, Gaussian mixture modeling of short-time Fourier transform features for audio fingerprinting, IEEE Trans. Inf. Forensics Secur. 1 (2006) 457–463.
[19] Y.H. Lin, M.S. Tsai, C.S. Chen, Applications of fuzzy classification with fuzzy C-means clustering and optimization strategies for load identification in NILM systems, in: IEEE Int. Conf. on Fuzzy Systems, 2011.
[20] N. Ramli, D. Mohamad, On the Jaccard index of similarity measure in ranking fuzzy numbers, Matematika 25 (2009) 157–165.
[21] J. Liang, S.K.K. Ng, G. Kendall, J.W.M. Cheng, Load signature study – Part II: Disaggregation framework, simulation, and applications, IEEE Trans. Power Deliv. 25 (2010) 561–569.
[22] S. Rahimi, A.D.C. Chan, R.A. Goubran, Usage monitoring of electrical devices in a smart home, in: IEEE Int. Conf. EMBS, September, 2011.
[23] R.A.S. Fernandes, I.N. Silva, M. Oleskovicz, Identification of residential load profile in the smart grid context, in: IEEE Power and Energy Society General Meeting, July, 2010.
[24] J.S.K. Leung, K.S.H. Ng, J.W.M. Cheng, Identifying appliances using load signatures and genetic algorithms, in: Int. Conf. of Electrical Engineering, July, 2007.
[25] H.H. Chang, L.S. Lin, N. Chen, W.J. Lee, Particle swarm optimization based nonintrusive demand monitoring and load identification in smart meters, in: IEEE IAS Annual Meeting, October, 2012.

[26] S. Rahimi, A.D.C. Chan, R.A. Goubran, Nonintrusive load monitoring id electrical devices in health smart homes, in: Proc. of IEEE I2MTC Conf., May, 2012.
[27] C. Belley, S. Gaboury, B. Bouchard, A. Bouzouane, An efficient and inexpensive method for activity recognition within a smart home based on load signatures of appliances, Pervasive Mob. Comput. 12 (2014) 58–78.
[28] M. Dong, P.C.M. Meira, W. Xu, C.Y. Chung, Non-intrusive signature extraction for major residential loads, IEEE Trans. Smart Grid 4 (2013) 1421–1430.
[29] S. Giri, M. Berges, A. Rowe, Towards automated appliance recognition using an EMF sensor in NILM platforms, Adv. Eng. Inform. 27 (2013) 477–485.